Amendments to the Claims

1. (Currently Amended) A system, comprising:

a psycho-physical state detection mechanism for detecting a psycho-physical state of a user based on ~~the~~ input speech data from the user; and

a spoken dialogue mechanism for carrying on a dialogue with ~~said~~ the user based on the psycho-physical state of the user, detected by the psycho-physical detection mechanism from the input speech data from the user.

2. (Currently Amended) The system according to claim 1, wherein ~~said~~ the spoken dialogue mechanism comprises:

a speech understanding mechanism for understanding the input speech data from the user based on the psycho-physical state of the user to generate a literal meaning of the input speech data; and

a voice response generation mechanism for generating a voice response to the user based on the literal meaning of the input speech data and the psycho-physical state of the user, wherein the voice response to the user is linguistically and acoustically adjusted according to the detected psycho-physical state of the user.

3. (Currently Amended) The system according to claim 2, wherein ~~said~~ the speech understanding mechanism comprises:

at least one acoustic model for characterizing ~~the~~ acoustic properties of the input speech data, each of ~~said~~ the at least one acoustic model corresponding to some distinct characteristic related to a psycho-physical state of a speaker;

an acoustic model selection mechanism for selecting an acoustic model that is appropriate ~~to~~ according to the psycho-physical state detected by the psycho-physical state detection mechanism;

a speech recognizer for generating a transcription of spoken words recognized from the input speech data using the acoustic model ~~selected~~ selected by the acoustic model selection mechanism; and

a language understanding mechanism for interpreting the literal meaning of the input speech data based on the transcription.

4. (Currently Amended) The system according to claim 2, wherein ~~said~~ the voice response generation mechanism comprises:

a natural language response generator for generating a response based on an understanding of the transcription, ~~said~~ the response being generated appropriately according to the psycho-physical state of the user;

a prosodic pattern determining mechanism for determining ~~the~~ a prosodic pattern to be applied to ~~said~~ the response that is considered as appropriate according to the psycho-physical state; and

a text-to-speech engine for synthesizing the voice response based on ~~said~~ the response and ~~said~~ the prosodic pattern.

5. (Currently Amended) The system according to claim 1, wherein ~~said~~ the psycho-physical state detection mechanism comprises:

an acoustic feature extractor for extracting acoustic features from the input speech data to generate at least one acoustic feature; and

a psycho-physical state classifier for classifying the input speech data into one or more psycho-physical states based on ~~said~~ the at least one acoustic feature.

6. (Original) The system according to claim 5, further comprising:

at least one psycho-physical state model, each of ~~said~~ the at least one psycho-physical state model corresponding to a single psycho-physical state and characterizing ~~the~~ acoustic properties of the single psycho-physical state; and

an off-line training mechanism for establishing ~~said~~ the at least one psycho-physical model based on labeled training speech data.

7. (Currently Amended) The system according to claim 1, further comprising a dialogue manager ~~that~~ to control the dialogue flow.

8. (Cancelled)

9. (Cancelled)

10. (Currently Amended) A method, comprising:

receiving, by a psycho-physical state detection mechanism, input speech data from a user;

detecting ~~the~~ a psycho-physical state of the user from the input speech data;

understanding, by a speech understanding mechanism, ~~the~~ a literal meaning of spoken words recognized from the input speech data based on the psycho-physical state of the user, detected by ~~said~~ the detecting; and

generating, by a voice response generation mechanism, a voice response to the user based on the literal meaning of the input speech data and the psycho-physical state

of the user, wherein the voice response to the user is linguistically and acoustically adjusted according to the detected psycho-physical state of the user.

11. (Currently Amended) The method according to claim 10, wherein ~~said~~ the detecting comprises:

extracting, by an acoustic feature extractor, at least one acoustic feature from the input speech data; and

classifying, by a psycho-physical state classifier and based on said at least one acoustic feature, the input speech data into the psycho-physical state according to at least one psycho-physical state model.

12. (Currently Amended) The method according to claim 11, further comprising:

receiving, by an off-line training mechanism, labeled training data, wherein each of the data items in ~~said~~ the labeled training data is labeled by a psycho-physical state; and

building ~~said~~ the at least one psycho-physical state model using the labeled training data, each of the at least one psycho-physical state model corresponding to a single psycho-physical state and being established based on the data items in the labeled training data that have a label corresponding to the single psycho-physical state.

13. (Currently Amended) The method according to claim 10, wherein ~~said~~ the understanding comprises:

selecting, by an acoustic model selection mechanism, an acoustic model, from at least one acoustic model, that is appropriate ~~to~~ according to the psycho-physical state, detected by ~~said~~ the detecting, each of ~~said~~ the at least one acoustic model corresponding to some distinct speech characteristic related to ~~a~~ the psycho-physical state;

recognizing, by a speech recognizer, the spoken words from the input speech data using the acoustic model, selected by ~~said~~ the selecting, to generate a transcription; and

interpreting, by a language understanding mechanism, the literal meaning of the spoken words based on the transcription.

14. (Currently Amended) The method according to claim 10, wherein ~~said~~ the generating comprises:

constructing, by a natural language response generator, a natural language response based on an understanding of the transcription, ~~said~~ the natural language response being constructed appropriately according to the psycho-physical state of the user;

determining, by a prosodic pattern determining mechanism, ~~the~~ a prosodic pattern to be applied to said natural ~~language~~ language response, wherein the prosodic pattern is considered to be appropriate according to the psycho-physical state; and

synthesizing, by a text-to-speech engine, the voice response based on ~~said~~ the natural language response and ~~said~~ the prosodic pattern.

15. (Cancelled)

16. (Cancelled)

17. (Cancelled)

18. (Currently Amended) A computer-readable medium encoded with a program, said program comprising instructions that when executed by a computer cause the computer to:

receive ~~receiving~~, by a psycho-physical state detection mechanism, input speech data from a user;

detect~~ing~~ ~~the~~ a psycho-physical state of the user from the input speech data;

understand~~ing~~, by a speech understanding mechanism, ~~the~~ a literal meaning of spoken words recognized from the input speech data based on the psycho-physical state of the user~~, detected by said detecting~~; and

generate ~~generating~~, by a voice response generation mechanism ~~mecahnism~~, a voice response to the user based on the literal meaning of the input speech data and the psycho-physical state of the user, wherein the voice response to the user is linguistically and acoustically adjusted according to the detected psycho-physical state of the user.

19. (Currently Amended) The medium according to claim 18, wherein ~~said~~ the to detect~~ing~~ comprises instructions that when executed by the computer cause the computer to:

extract~~ing~~, by a acoustic feature extractor, at least one acoustic feature from the input speech data; and

classify~~ing~~, by a psycho-physical state classifier and based on ~~said~~ the at least one feature, the input speech data into the psycho-physical state according to at least one psycho-physical state model.

20. (Currently Amended) The medium according to claim 19, further comprising instructions that when executed by the computer cause the computer to:

receive ~~receiving~~, by an off-line training mechanism, labeled training data, wherein each of the data items in ~~said~~ the labeled training data is labeled by a psycho-physical state; and

building ~~said~~ the at least one psycho-physical state model using the labeled training data, each of the at least one psycho-physical state model corresponding to a single psycho-physical state and being established based on the data items in the labeled training data that have a label corresponding to the single psycho-physical state.

21. (Currently Amended) The medium according to claim 18, wherein ~~said~~ the understanding comprises instructions that when executed by the computer cause the computer to:

select~~ing~~, by an acoustic model selection mechanism, an acoustic model, from at least one acoustic model, that is appropriate ~~to~~ according to the psycho-physical state, ~~detected by said detecting,~~ each of ~~said~~ the at least one acoustic model corresponding to some distinct speech characteristic related to a psycho-physical state;

recognize ~~recognizing~~, by a speech recognizer, the spoken words from the input speech data using the acoustic model, ~~selected by said selecting,~~ to generate a transcription; and

interpret ~~interpreting~~, by a language understanding mechanism, the literal meaning of the spoken words based on the transcription.

22. (Currently Amended) The medium according to claim 18, wherein ~~said~~ the to generate ~~generating~~ comprises instructions that when executed by the computer cause the computer to:

construct~~ing~~, by a natural language response generator, a natural language response based on an understanding of the transcription, ~~said~~ the natural language response being constructed appropriately according to the psycho-physical state of the user;

determine ~~determining~~, by a prosodic pattern determining mechanism, ~~the~~ a prosodic pattern to be applied to ~~said~~ the natural ~~language~~ language response, wherein the prosodic pattern is considered to be appropriate according to the psycho-physical state; and

synthesize ~~synthesizing~~, by a text-to-speech engine, the voice response based on ~~said~~ the natural language response and ~~said~~ the prosodic pattern.

23. (Cancelled)

24. (Cancelled)